ANALYSIS OF AIRPORT GROUND DELAY PROGRAM DECISIONS USING DATA MINING TECHNIQUES

Deepak Kulkarni, Yao Wang and Banavar Sridhar

NASA Ames Research Center, Moffett Field, CA

Abstract

Air traffic service providers have to make decisions regarding changes to air traffic flow in the event of major weather disturbances and traffic congestions to maintain safety of the system. The behavior of the air traffic management system will be more predictable if consistent decisions are made under similar traffic and weather conditions. Consistency of deciding on control action depends on the weather and traffic conditions as well as accuracy in predicting these conditions. Weather parameters (defined in terms of forecast and actual weather and traffic conditions) on different days can be used to categorize days into days with little decision consistency, days with moderate decision consistency and days with high decision consistency. Four years of traffic, weather and ground delay program decisions data at major airports in the United States are used in the analysis. This paper examines performance of different data mining methods in the three regions of decision consistency. Not surprisingly, data mining methods have the best performance in the region of most decision consistency and have the poorest performance in the region of little decision consistency. In applications where data mining methods have differing performance in differing regions, it would be more useful to characterize the region specific performance instead of characterizing performance by a single parameter. Finally, the results show no significant variation in the performance of different data mining methods for this particular problem. The fact that different mining methods show no significant variation also provides further confidence in the results of data mining methods. Work in this abstract discusses initial results. This paper describes the results in terms of both forecast and actual environmental conditions and discusses how prediction errors impact decision consistency.

Section 1. Introduction

The continuous growth in the demand for air transportation results in an imbalance between airport capacity, airspace capacity and traffic demand. Airport arrival rates and airport departure rates are affected by weather conditions. The airspace capacity of a region depends on the ability of the system to maintain safe separation between aircraft in the region. The airspace capacity is severely limited by inclement weather. FAA has a national center called Air Traffic Control Similarly, airport arrival rates and airport departure rates are reduced in the presence of bad weather. System Command Center (ATCSCC) that oversees national traffic. Traffic managers at Air Route Traffic Control Center (ARTCC) collaborate with dispatchers at various Airlines' Operations Center (AOC) to mitigate the demand-capacity imbalance caused by weather. The end result is the implementation of a set of Traffic Flow Management (TFM) initiatives such as ground delay programs, reroute advisories, flow metering, and ground stops.

Data Mining is the automated process of analyzing large sets of data and then extracting patterns in the data. Data mining tools are capable of predicting behaviors and future trends, allowing an organization to

benefit from past experience in making knowledge-driven decisions.

In recent years, a number of GDP-related studies using data-mining algorithms have appeared in the literature [1]. Since GDP operations are largely developed and carried out with limited decision support tools in current operations, techniques for modeling the impact of GDP programs prior to operational implementation have been researched in recent years. In [2], a decision support capability to predict Aircraft Arrival Rates (AAR) and to determine Ground Delay Program (GDP) program rate and duration based on Terminal Aerodrome Forecast (TAF) weather forecast data using Support Vector Machine (SVM) algorithm, is described. The uses of Ensemble Bagging Decision Tree (BDT), SVM, or Neural Networks (NN) methods to predict the airport capacity and GDP parameters with weather and airport data are introduced in [3] and [4]. Despite the past work in this area, there are no published systematic studies seeking to evaluate and predict whether a GDP operation is required or not for days having similar weather and airport conditions.

Data with past information about traffic and weather conditions as well as traffic management initiatives has been archived in ASPM and NTML databases. However, such databases do not provide information in a form that can be used in decision making. Data mining algorithms have the potential to develop associations between weather patterns and the corresponding ground delay program responses. If successful, they can be used to improve and standardize TFM decisions resulting in better management of traffic flows on days with reliable weather forecasts. The approach here seeks to develop a set of data mining and machine learning models and apply them to historical archives of weather observations and TFM initiatives to determine the extent to which the theory can predict and explain the observed traffic flow behaviors.

In this study, the major sources of data that were used include: the National Traffic Management Log (NTML) and Aviation System Performance Metrics (ASPM). The data used was from the years 2006 to 2010. The NTML is a unified system developed by the FAA that is used to automate coordination, logging and communication of traffic management initiatives in the NAS. For the purpose of this initial study, the GDP entries in NTML were used as inputs to the data mining algorithms.

A brief overview of the remainder of the paper is as follows. Section 2 discusses ground delay programs. Section 3 provides a high-level overview of data mining techniques that were employed in this study. Section 4 describes the methodology used in the program including metrics and data used in the study. Section 5 presents results. Section 6 provides a conclusion based on the results.

Section 2. Statistics of Ground Delay Programs

The mission of the FAA's traffic management system to balance traffic demand with system capacity is achieved through a variety of Traffic Management Initiatives (TMI) instituted and modified by traffic managers at the regional and national levels. The FAA developed the National Traffic Management Log (NTML) to provide a single system for automated coordination, logging, and communication of TMIs throughout the National Airspace System. Figures below show more detailed GDP event statistics from the data.
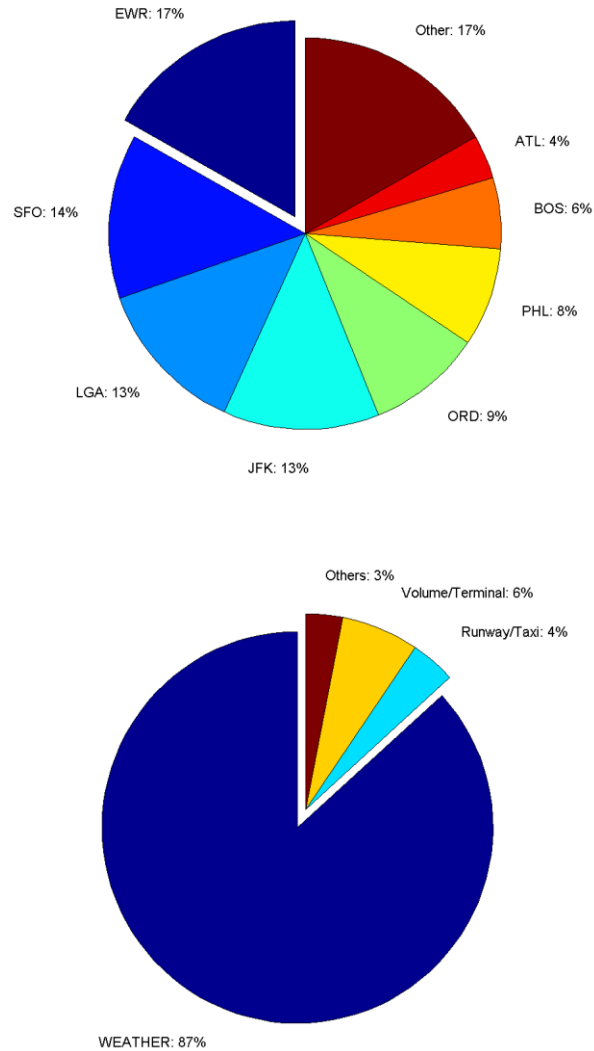
Figure 1: The ratios of GDP counts and GDP causes for the top 8 US airports

Figure 1 displays the ratios in percentage between airport GDP counts and the total NAS GDP counts for the top 8 airports. Figure 1 shows that the most frequent demand-capacity imbalances occurred at the airports in the northeast region of the United States, such as the three New York-area airports (EWR, LGA, and JFK), Philadelphia (PHL), and Boston Logan International Airport (BOS). The major cause of Ground Delay Programs is weather as demonstrated in Fig.1. The diverse weather subcategory causes are presented in Figure 2. Altogether, these data illustrate that the dominated weather causes for GDPs are different at different airports. For example, while close to 90% low ceilings due to marine stratus cause GDPs at SFO, wind accounts for about 50% of GDPs at the three New York-area airports, and thunder storms are the major sources of GDP at ATL.
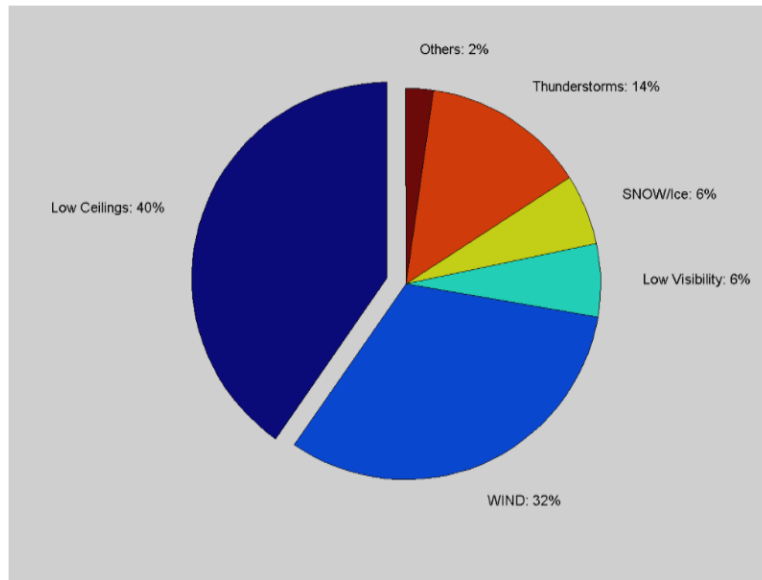
Figure 2: Ratios of the counts between weather subcategories and the total weather GDPs

As Newark international airport is an airport that has a very high number of ground delay programs and that contributes significantly to national airspace delays, we initially focused this study at this airport. ASPM is an FAA database containing airport specific data, such as throughput and the weather impacting the airport. Hourly values of wind speed, visibility, ceiling, Instrument Meteorological Conditions (IMC), scheduled arrivals and departures from ASPM data were used to compute input variables in this study.

IMC impacted traffic and wind impacted traffic are two parameters derived from traffic and weather data. As weather impact on the national airspace depends on how many aircraft are impacted by inclement weather, we are using these two metrics to capture the impact of weather on traffic. Wind impacted traffic at an airport was defined as the number of arriving or departing aircraft while wind speed is over 15 knots. Similarly, IMC impacted traffic at an airport was defined as the number of arriving or departing aircraft while there are IMC conditions. Figure 3 below shows histograms of IMC WITI and wind WITI values over at EWR a period of five years. Mean AAR and ADR at EWR during the period of this study was 41. 26% of days have IMC conditions for five hours or more and 30% of the days have average wind speed of 10 knots or more.

Daily values of wind WITI and IMC WITI , wind speed, visibility, and ceiling were computed as daily average of hourly values. Values for daily variation in visibility and ceiling are computed as difference between daily maximum and daily minimum values of visibility and ceiling. Daily weather and traffic signature was characterized with following parameters: wind speed, variation in wind speed, visibility, variation in visibility, ceiling, variation in ceiling, Instrument Meteorological Conditions (IMC), scheduled arrivals, IMC impacted traffic and wind impacted traffic.

Principle components analysis of these 10 variables found that the most relevant variables are IMC impacted traffic and wind impacted traffic.
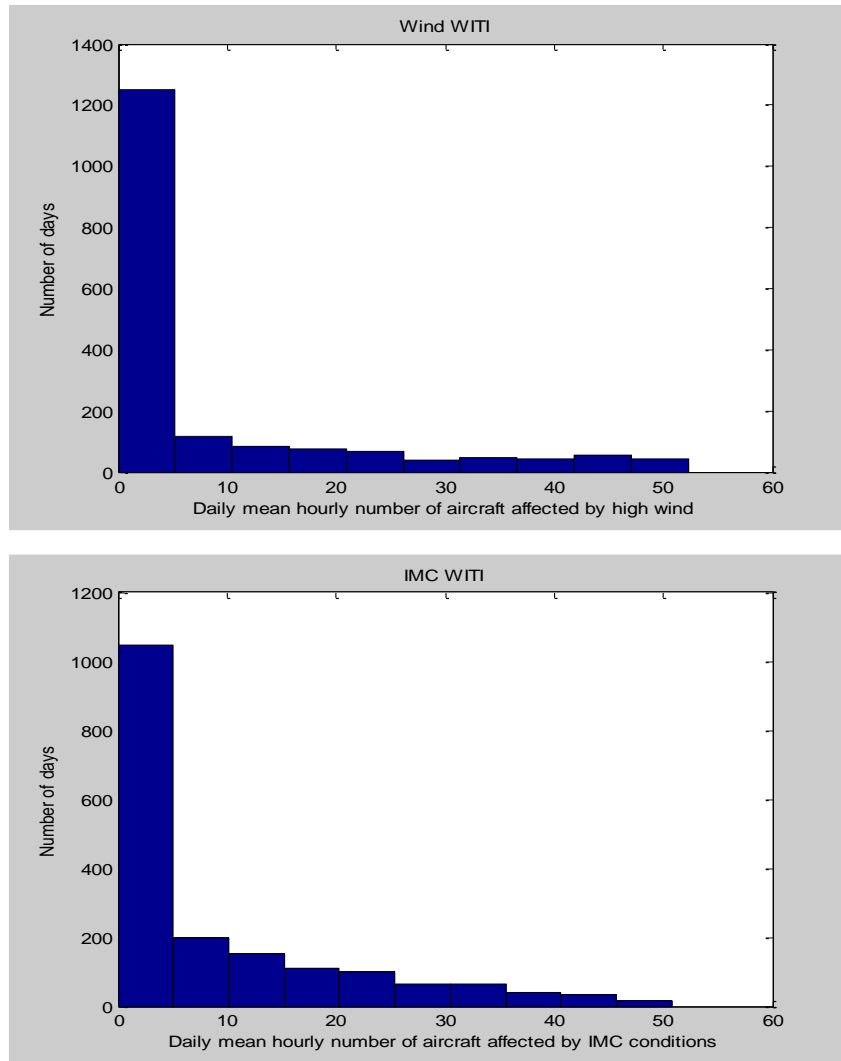
Figure 3. Histograms of IMC WITI and Wind WITI

Section 3.   Data Mining Overview

In our study, we use three data mining methods: ensemble bagging decision trees (BDT), nearest neighbor classifier and support vector machine (SVM) learning algorithms. These models were selected because decision surfaces modeled by these are very different from each other. As we do not have direct knowledge of shape of decision surface in the case of GDP decisions, it would be good to use different data mining methods to examine if the actual surface is easier modeled by one of these methods.

A. Decision Tree Classifier:

Decision tree learning (Quinlan, 1986) uses a decision tree as a predictive model that associates input variables with target values. Each internal node corresponds to a condition on an input variable; there are edges from the node to children for each of the possible values of that input variable. Each leaf node a

value of the target variable associated with it. This value is the predicted value given the values of the input variables represented by the path from the root of the tree to the leaf. Algorithm for creating decision trees works top-down by selecting a condition on a variable at each step that best splits the set of items. Algorithm used in this study used a metric called "Gini impurity."   Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly classified if it were randomly classified according to the distribution of classes in the subset. Another commonly used measure is "information gain" measure.

B. Nearest Neighbor Classifier:

The Nearest Neighbor classifier (Devroye, 1986) involves a training set of both positive and negative cases. A new sample is classified by calculating the distance to the nearest training case; the class associated with that point then determines the classification of the sample. The k-NN classifier extends this concept by examining the k nearest points and assigning the class of the majority. It is common to select k small and odd to break ties (typically 1, 3 or 5). Larger k values help reduce the effects of noisy points within the training data set, and the choice of k is often performed through cross-validation.

C. Support Vector Machine (SVM)

A support vector machine (Cortes, 1995) constructs a hyper-plane or set of hyper-planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Its robust performance with respect to limited, sparse and noisy data is making it widely used in many applications from protein function, and face recognition, to text categorization for classification and regression prediction. The SVM model has also been utilized in airport capacity classification prediction.

When used for binary classification, the SVM algorithm separates a given set of two-class training data by constructing a multidimensional hyper-plane that optimally discriminates between the two clusters. Although SVMs were originally proposed to solve linear classification problems, they can be applied to non-linear decision functions by using the so-called kernel function trick. Adopting this kernel technique, SVM can be utilized to automatically realize a non-linear mapping to a high dimensional space. The hyper plane in the high dimensional space corresponds to a non-linear decision boundary in the input space. A widely used kernel is the Gaussian radial basis function (RBF).

Section 4. Methodology

In some applications, different operators may take different control actions in the presence of similar weather and traffic conditions. Sometimes, the same operator may take different control actions in the presence of similar weather conditions.   The reasons for this may be various. Inconsistency may be owing to differing objectives, decision-making styles, or training. The degree of operator decision consistency varies in different regions of the state space. It can be useful to understand the nature of decision inconsistency.

Furthermore, the performance of these data mining methods will vary depending on the state of the system as specified by the observations. The ability of machine learning depends on the consistency of the decision-making process and the availability of the training data in the various regions of the input data state space. Another factor complicating the analysis is lack of clear criterion driving the control actions resulting in different decisions for the same values of the state space. Given the variability in the

performance of data mining methods, using a single number to characterize predictive accuracy is not helpful. The paper discusses how to divide data into regions with differing decision consistency and report performance of different data mining methods in the different regions of decision consistency. We will also examine if there is variation in the performance of different data mining methods.

Approach

The general approach adopted in this learning automation work involves the following steps.

1 Division of data into regions of differing decision consistency

2 Comparison of performance of two data mining methods in the regions of differing decision consistency

3 Analysis of sensitivity of results to weather prediction inaccuracies

Section 5. Results

Regions of differing decision consistency
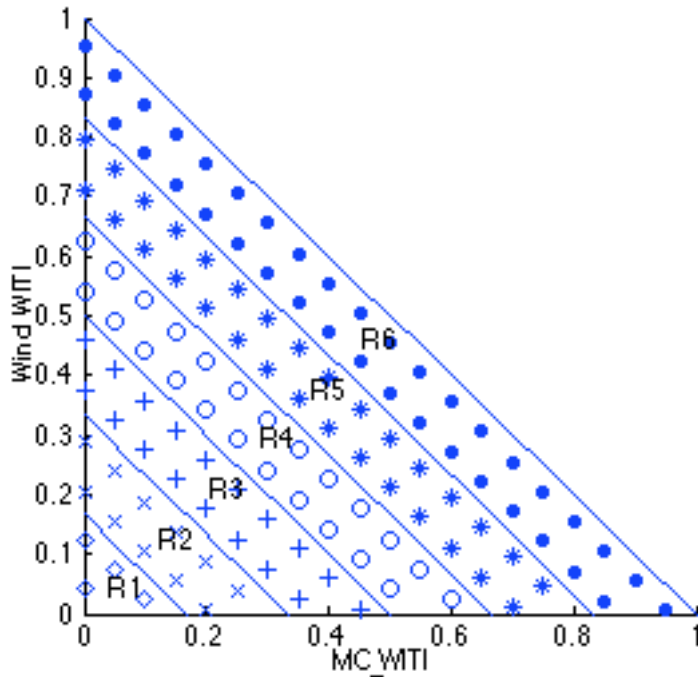
For the purpose of analysis, we divide the data into six regions as shown in the figure above. Division is based on the values of sum of Wind WITI and MC WITI. R1 is a region where the sum of MC WITI and Wind WITI is the lowest. R6 is a region where the sum is the highest.

Percentile values of sum of MC WITI and Wind WITI	Probability of GDP occurrence	Decision consistency	Percent of data
R1	.20	.80 High	25
R2	.26	.74 Medium	8
R3	.42	.58 Low	17
R4	.66	.66 Low	17
R5	.79	.79 Medium	16
R6	.95	.95 High	17

Table 1: Segmentation of Data into Multiple Regions

Difficulty of deciding on control action depends on the region of variable space. For example, on clear weather days, most operators would not have any difficulty in concluding that there is no need for weather-caused GDP. Similarly, on really bad weather day, most operators would conclude that there is a need for weather-caused GDP.

Table 1 shows segmentation of data into 6 regions depending on the sum of Wind WITI and MC WITI and the probability of GDP occurrence in this region of data. Low is defined as [.5, .66], medium is defined as

[.67, .79] and high as defined as [.8,1.0]. The number shown in the second column of the table is the percent of cases with GDPs in the particular region of interest. Decision consistency refers to percent of days when the decision was in agreement with the majority decision for the region. Figure 4 shows GDP decisions and weather conditions for all days analyzed in this study. Each point in the plot corresponds to a day with IMC and wind values corresponding to it. The point is represented by "x" if there is a GDP on that day. It is represented by "o" if there is no GDP on that day. In the figure 4, the top, middle and bottom plots show the regions of low, medium and high decision consistency, respectively. As evident in the table 1 and figure 4, this value depends on the region of variable space. For example, the first row in the table corresponds to mostly clear weather days. In this case, most operators do not have any difficulty in concluding that there is no need of weather-caused GDP.
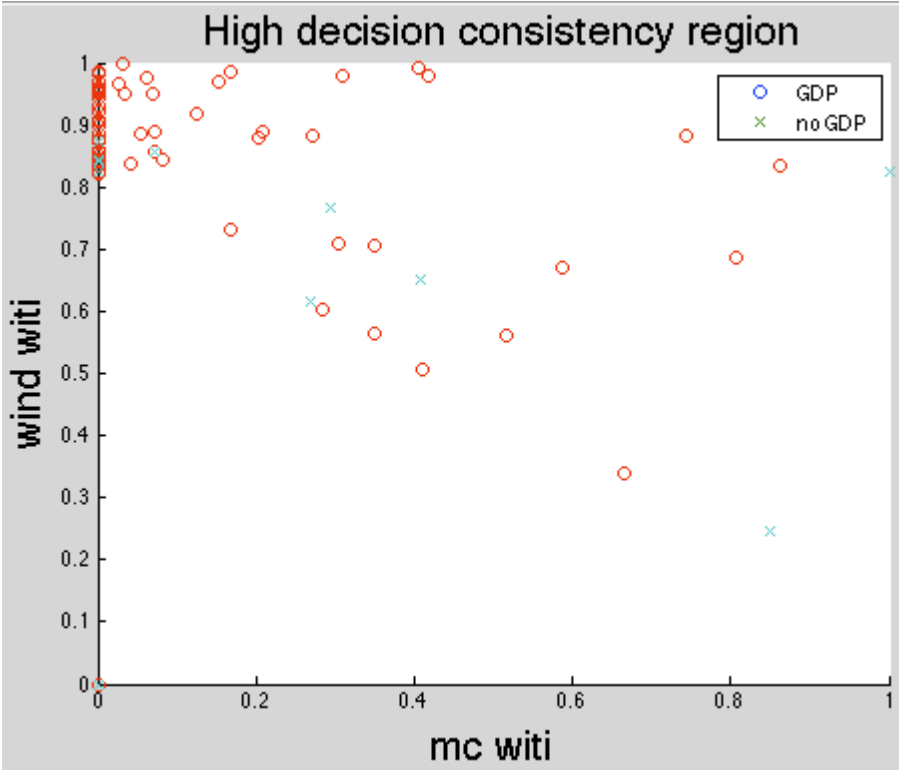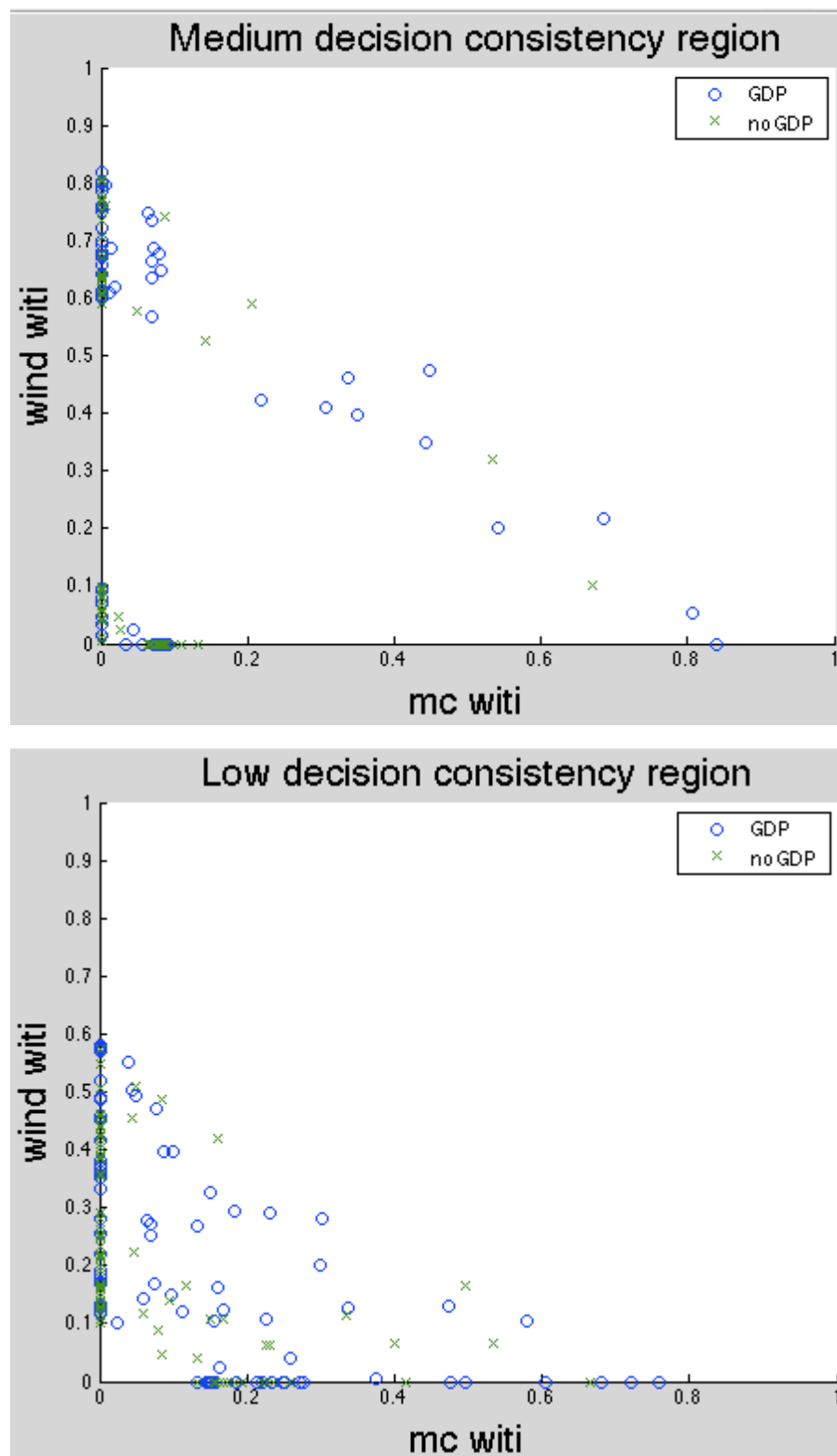
Figure 4: Differing Decision Consistency in Different Regions

On the other hand, the last row in the table corresponds to days with the worst weather. In this case, 95% of operators concluded that there is a need of weather-caused GDP. If we examine third row in the table, we find that about 58% of operator chose to implement GDPs and 42% chose not to. Operators probably need a decision support system in the cases where there seem to be divergence of control actions under the exact same conditions. Given the divergent characteristics of different regions, it would be useful to examine the performance of data mining methods in different regions of data space.

We categorized the six regions as having low, medium or high level of decision consistency and then compared performance of different methods when data has these differing levels of decision consistency.

We found that about 34% of days fall in the category of low decision consistency. About 24% fall in the category of moderate decision consistency and about 42% of days fall in the category of high decision consistency. We also find that performance of data mining methods is better in the region of high decision consistency and is poorer in the region of low decision consistency.

Performance of different methods in the regions of differing decision consistency
Accuracy of these methods varies depending on region of decision consistency. For example, decision tree classifier had overall accuracy of 82% in the region of high decision consistency, an accuracy of 69% in the region of medium decision consistency and an accuracy of 59% in the region of low decision consistency. This is not surprising as data mining models can only be as good as the data on which they are trained on.

Utility of data mining methods may vary in different regions of decision consistency. There is probably no need for data mining assistant system in the region of high decision consistency. Data mining methods can be useful in the regions of medium and low decision consistency, but their accuracy is the lowest in the region of low decision consistency.

Decision Consistency Region	% data	Nearest Neighbor Classification	SVM	Decision Tree Classification
Low	34	58%	61%	59%
Medium	24	63%	55%	69%
High	42	78%	75%	82%

Table 2: Data Mining Method Performance:

Relevant Inaccuracies in weather predictions

Natural approach is using an arbitrary threshold in deciding if weather prediction was inaccurate. For example, if difference in predicted wind speed and actual wind speed is 5 knots, prediction was inaccurate. However, whether wind speed is 0 or 5 may not affect control decisions. Similarly, whether wind speed is 30 or 35 maybe equally bad.

Our definition attempts to capture inaccuracies in weather prediction that are likely to have impact on control decisions. We create a decision tree predicting control decisions based on historical METAR data. Then, we use this tree to predict control decision first using available METAAR data. Then, we use the same tree to predict control decisions using available TAF data. If these two decisions are different, then we regard the weather prediction to be inaccurate. There were about 30% of days with significant weather accuracies with this definition.

Sensitivity of results to weather prediction accuracy

Data segmentation method described in the previous section is not the only method by which we could divide the data. In this subsection, we examine the sensitivity of our conclusions to the method used to weather prediction accuracy. To do so, we divide the data into days when weather prediction is accurate and into days when weather prediction is not accurate. Table 3 shows that differing region of decision inconsistency exist even on days with correct weather prediction. Table 4 shows that there is a very high level of decision uncertainty on the days when weather predictions are inaccurate.

Range of values of sum of MC WITI and Wind WITI	Probability of GDP occurrence	Decision consistency	% data
R1	.15	.85 High	26
R2	.21	.79 Med	7
R3	.44	.56 Low	17
R4	.85	.85 High	17
R5	.83	.83 High	16
R6	.98	.98 High	17

Table 3: Segmentation of Days with Accurate Weather Prediction

Range of values of sum of MC WITI and Wind WITI	Probability of GDP occurrence	Decision consistency	% data

R1		.33	.67 Low	24
R2		.42	.56 Low	9
R3		.42	.58 Low	17
R4		.56	.56 Low	17
R5		.52	.52 Low	16
R6		.66	.66 Low	17

Table 4: Segmentation of Days with Inaccurate Weather Prediction

Section 6. Conclusion

Difficulty of deciding on control action depends on the region of variable space. Weather and traffic signature on different days can be used to categorize days into days with little decision consistency, days with moderate decision consistency and days with high decision consistency. This paper reported performance of different data mining methods in the three regions of decision consistency.   Not surprisingly, data mining methods have the best performance in the region of high decision consistency and have the poorest performance in the region of low decision consistency. In applications where data mining methods have differing performance in differing regions, it would be more useful to characterize the region specific performance instead of characterizing performance by a single parameter.

Also, there is probably not need for data mining assistant system in the region of high decision consistency.   Therefore, operators may find decision support systems to be most useful in the regions of moderate or low decision consistency. Also, organizations may want to examine decision-making processes that are used in these regions to see how much subjectivity exists.   Thus, it may be useful to segment data and identify the regions of low and moderate decision consistency.

Finally, we also found that there was not significant variation in the performance of different data mining methods for this particular problem. The fact that different mining methods show no significant variation also provides further confidence in the results of data mining methods.

We also found that differing region of decision consistency exist even on days with correct weather prediction. On the other hand, there is a low level of decision consistency on the days when weather predictions are inaccurate.

Bibliography

[1] Cortes, C.; Vapnik, V. (1995). "Support-vector networks". Machine Learning 20 (3): 273

[2] Devroye, L., Gyorfi, L. & Lugosi, G. (1996). A probabilistic theory of pattern recognition. Springer. ISBN 0-3879-4618-7.

[3] Klein, Alexander, S. Kavoussi, and R. S. Lee. "Weather Forecast Accuracy: Study of Impact on Airport Capacity and Estimation of Avoidable Costs,." *Eighth USA/Europe Air Traffic Management Research and Development Seminar .* 2009.

[4] Quinlan, J. R., (1986). Induction of Decision Trees. Machine Learning 1: 81-106, Kluwer Academic Publishers

[5] Smith, David A., L. Sherry, G. Donohue. "Decision Support Tool for Predicting Aircraft Arrival Rates, Ground Delay Programs, and Airport Delays from Weather Forecasts, ." *Proceedings International Conference on Research in Air Transportation.* Fairfax, VA. , 2008.

[6] Wang, Yao. "Prediction of weather impacted airport capacity using ensemble learning,." *Digital Avionics System Conference.* 2011.

[7] Wang, Yao, and D. Kulkarni. "Modeling Weather Impact on Ground Delay Programs." *SAE International Journal of Aerospace* 4, no. 2 (November 2011): 1207-1215.